Part 17: Linear Basis Functions

March 29, 2021

This should be a fun easy little jaunt into the world of linear basis functions, which will be based heavily on Martin Krasser's post regarding the same subject.

1 Linear Basis Function Concept

The idea starts with the basic assumption of any statistical model, with the **model structure**:

$$f(x) = w^T x \tag{1}$$

With the **statistical structure**:

$$y = f(x) + \epsilon \tag{2}$$

Which can be combined into a **statistical model structure**:

$$Y \sim N(f(x), \beta^{-1})$$

Where we are assuming the model $f(x)$ is distributed normally with a mean (given by the model structure) and variance β^{-1} for inputs x and outputs y. Like all distributions this has a likelihood $L(x) = \Pi_i^N N(f(x_i), \beta^{-1})$ and like all normal distributions has an analytical likelihood (that we will not discuss because we've gone over it before).

2 Making it Bayesian

How can we make this Bayesian? Of course we need a prior. The prior we will use is:

$$w \sim N(0, \alpha^{-1}I)$$

Which will constrain w in the hyperparameter optimization around 0 with some variance α^{-1}. The MLE estimate for the hyperparameters w (or rather MAP estimate because it has a prior) is as follows:

$$\hat{w} = \beta(\alpha + \beta\Phi\Phi^T)^{-1}\Phi^T Y$$

$$\hat{\Sigma}_w = (\alpha + \beta\Phi\Phi^T)^{-1}$$

Which is also normally distributed. Note that $\Phi(x)$ is the design matrix for all inputs x, which forms the backbone of the basis function. So if we have a linear basis function we have $\phi(x) = [1, x]$ and for a second order polynomial we have $\phi(x) = [1, x, x^2]$ and so on and so forth. The capital letter $\Phi(x)$ just indicates this matrix calculation for all x in a group.

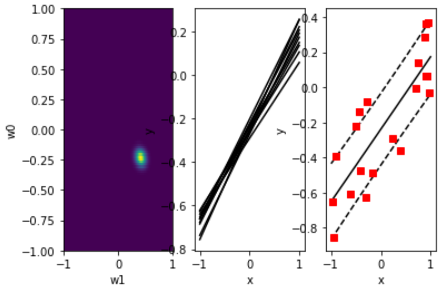
3 Regression

The most important part is of course regression! Using the optimal MLE hyperparameters we get a mean and covariance function (to take the standard deviation at a single point we take the sum of a column of this covariance matrix square rooted):

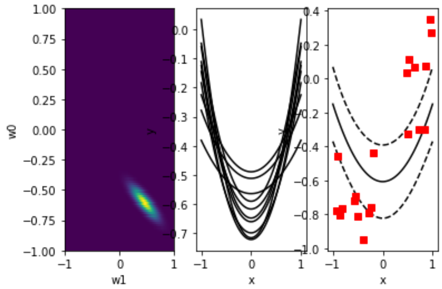
$$\hat{y} = \hat{w}^T x \tag{3}$$

$$\hat{\Sigma}_y = \phi(x)^T \hat{\Sigma}_w \phi(x) + \beta^{-1} \tag{4}$$

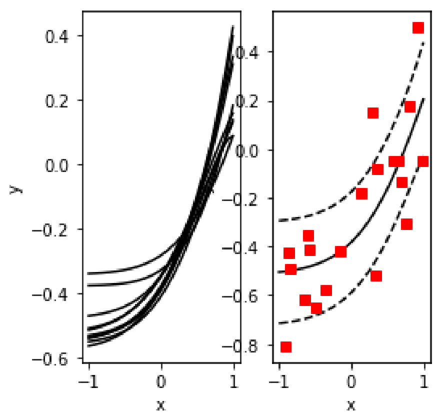
As an example we generate data X and Y for a linear problem with normally distributed errors and get the following:



(a)



(b)



(c)

Figure 1: (a) $[1, x]$ (b) $[1, x^2]$, (c) $[1, K_{SE}(x, \mu = 2, \sigma = 0.1)]$